THE REALITY OF SCIENCE WITH NO SECRETS

Amelia Hunt

School of Biological Sciences,
THE UNIVERSITY OF EDINBURGH,
Edinburgh, UK.

ORCID: 0000-0001-9953-9556

Undergraduate Honours Research Project Report,
School of Biological Sciences, THE UNIVERSITY OF EDINBURGH, Edinburgh, UK
Project Supervisor: Andrew J. Millar

Table of Contents

Abstract

Open Notebook Science consists of putting one's whole research notebook freely online, in real time. It is at the extreme end of the global movement towards greater transparency and accessibility in scientific research. Whilst many organisations and countries are supporting the Open Science movement, it seems that Open Notebook Science has failed to gain traction. Therefore, this project aimed to understand experiences of Open Notebook Science and to address a gap in formal evidence. An extensive literature review to decide the scope of the project resulted in qualitative, semi-structured interviews of five, international advocates of Open Notebook Science. Analysis of the interviews yielded four themes, revealing both positive and negatives outcomes from practising Open Notebook Science. The results suggest that Open Notebook Science is unlikely to be widely adopted in the future, given the intentional extremism of the approach, the negative experiences discovered and the current requirements of scientific disciplines. Specific, under-served research areas may benefit from this approach. However, promoting a range of openness, including different levels of Open notebooks, may be both beneficial and achievable. Open Notebook Science is a trailblazing contribution to that goal. Future research should aim to discover a wider range of opinions and discover the prevalence of Open Notebook Science more directly.

1. Introduction

1.1 Open science

Experimental data is the backbone of scientific research. Yet, how it is used, displayed and accessed is just as vital as the results themselves. Openness, reproducibility and transparency are fundamental aspects for the scientific community to gain trust and confidence in results and, recently, more attention has been paid to how scientific data is published and used (McNutt, 2014; Nosek *et al.,* 2015). The Royal Society (2012) highlighted the importance of openness in science, but also demonstrated the global movement towards more open research. For example, the UK Concordat on Open Research Data (2016) stated that we are on target to have all tax-funded publications open to the public and highlights benefits of openness, stating that having data available for analysis and reuse can increase public trust, resource efficiency, and economic growth. This demand for transparency has resulted in numerous ways of uploading and storing data as well as allowing collaboration and communication not previously available (Insel, 2013; Royal Society, 2012). Traditionally, focus has remained on publishing work in peer-reviewed journals (De Roure *et al.,* 2010). However, with widespread internet use and the scientific journal being over 350 years old, it is unsurprising that new forms of sharing have emerged (De Roure *et al.,* 2010; Holmberg *et al.,* 2015; Piwowar and Chapman, 2008). This push towards greater openness has led to the Open Science movement (Grand *et al.,* 2016; Wald, 2010). Open Science encompasses different concepts including Open Data, Open Access, Open Source and Open Notebook Science (ONS) (Wald, 2010). Open Data refers to making data free and available without copyright restrictions, whilst Open Access describes making scholarly and peer-reviewed journal articles freely available, and Open Source refers to software code that is available to be reused and modified (Wald, 2010). Overall, the movement intends to make scientific research more accessible, efficient, and cooperative (Friesike and Schildhauer, 2014; Grand *et al.,* 2016). Whilst all the Open Science practises aim to achieve this, ONS is the most extreme.

1.2 History and Definition of Open Notebook Science (ONS)

ONS is “making the entirety of one’s laboratory notebook and all associated raw data public in as close to real time as possible” (Stafford, 2010). It involves publishing all data, including information not normally disclosed such as failed experiments, online, freely allowing public access. The term ‘Open Notebook Science’ was coined by Jean-Claude Bradley in 2006, with the primary philosophy being “no insider information” (Poynder, 2010). Bradley created ONS after realising that his research was not having the impact that he wanted, and that this was partly due to secrecy. In 2008 he created the ONS Challenge. This encouraged students to use open-notebooks when discovering the solubility of organic compounds, and allowed him to reach out to the wider scientific community (Bradley *et al.,* 2010; Poynder, 2010). Whilst ONS initially received interest, it also split opinions.

1.3 Arguments For and Against ONS

Arguments for ONS include that it allows science to move faster due to ease of collaboration such that others can solve problems or spot mistakes in the data (Gewin, 2013; Stafford, 2010). Also, Bradley claimed that the ability to see failed experiments is vital (Poynder, 2010). This value is currently overlooked as there is a large publication bias in various disciplines towards statistically significant results, versus studies which produce null results (Franco *et al.,* 2014). Nevertheless, drawbacks of ONS have also been suggested. The main concerns are theft, or scooping, and the inability to publish work due to an open notebook being considered a pre-print (Wald, 2010). Despite some arguments being documented, evidence in the literature is scarce.

1.4 Importance of this Study

It is beneficial to investigate ONS for numerous reasons. Firstly, despite the clear movement towards greater transparency and accessibility, it seems ONS has failed to gain widespread support (Friesike and Schildhauer, 2014). With over 10 years since ONS was devised, a better idea of its prevalence would be valuable (Murray-Rust, 2014). Therefore, a study on ONS is required to gain more understanding. Secondly, there is no systematic evidence on ONS. Literature specifically on ONS is confined to short, casual interviews with Bradley or occasionally another practitioner, articles explaining the practise and giving the author's opinion, and blogs of ONS practitioners (Bergman, 2014; Clinio and Albagli, 2017; Gewin, 2013; Neylon, 2007a; Poynder, 2010; Wald, 2010). However, these interviews use no formal methodology or analysis, only providing snippets of information from one individual. Meanwhile, there is more evidence on other parts of Open Science. For example, the Wellcome Trust commissioned a survey that gathered evidence on researchers' attitudes and behaviour towards open research (Van den Eynden *et al.,* 2016). Finally, scientific data is an invaluable resource, so how it is used and accessed must be carefully determined. So, it is vital to consider all components of Open Science, including ONS, to help answer the larger question of how open science should be (Randles *et al.,* 2017). Overall, this study is addressing a major gap in literature by gathering evidence on ONS, which will inform debates on whether ONS should be more widely considered.

Due to the lack of precedent studies, a key part was to determine the best method to gain an insight into ONS today. So, the scope and specific objectives of this project were determined after an extensive literature review on ONS, Open Science, and methodological papers. Ultimately, we have little understanding of ONS in practice, so discovering experiences of ONS practitioners is valuable research. It was determined that qualitative interviews of advocates of ONS would provide an information-rich insight into ONS and shed light on its future potential. This is further discussed in Results.

1.5 Research Aims and Objectives

The aim of this project was to provide the first systematic evidence on ONS. Specifically, to discover experiences and opinions of advocates of ONS, in relation to their time supporting and practising ONS. Based on the reviewed literature, the objectives were:

1. Determine an appropriate experimental design to investigate ONS today
2. Become familiar with qualitative interview techniques to determine the most suitable sampling and analysis methods
3. Discover any common opinions and experiences of advocates of ONS
4. Use the interview data to determine the prevalence of ONS
5. Draw conclusions about how ONS may be used in the future and provide a relevant basis for further research

2. Results

2.1 Experiment design

An extensive literature review was conducted to determine the most suitable experimental protocol for investigating ONS. A qualitative semi-structured interview approach was chosen as it allowed in-depth understanding of the topic and detailed accounts from a few selected individuals (DiCicco-Bloom and Crabtree, 2006; Whiting, 2013). Meetings with an expert in qualitative research confirmed that my methodology and interview guide *(Appendix A)* were suitable (please see Materials and Methods for further details). The sample was chosen to be advocates of ONS as literature review revealed that ONS is a niche activity, with few practitioners. Therefore, advocates likely have significant familiarity with ONS, providing information-rich interviews (Padoan *et al.,* 2016). Another group of interviewees was not used alongside because the objective was to learn about a specific subgroup in-depth, not compare results between strata at this stage (Palinkas *et al.,* 2013). A questionnaire to discover the attitudes of many people was considered. However, limitations include only having set responses due to its quantitative nature, so less detailed information on ONS would be acquired (Palinkas *et al.,* 2013). Given the timing and resource constraints, it would not have been possible to gather global Biological Scientists' opinions. It would likely be limited to the University of Edinburgh Biological Sciences department, which may have response biases due to University policies. Therefore, I decided that qualitative interviews would provide the most valuable insight.

2.2 Sample statistics

Participants were selected for interview via purposeful sampling, by which individuals are selected non-randomly based on characteristics they possess. Specifically, mixed-method homogeneity and snowball sampling were used. Homogeneity sampling is selecting a specific group of individuals due to their particular qualities and experiences (Etikan, 2016). Snowball sampling involves identifying participants from a source individual who recommends those with similar characteristics (Palinkas *et al.,* 2013). Snowball sampling was used because it is useful for finding small or difficult to reach communities, and increases the chance of people agreeing to interview, due to trust between participants (Sadler *et al.,* 2010). Also, it permits access to naturally interacting populations (Biernacki and Waldorf, 1981; Coleman, 1958). I initially aimed to investigate ONS within Biological Sciences. However, literature review and snowball sampling results revealed the sample size would be too small (figure 1). Therefore, potential participants from any scientific discipline were contacted. In total, five interviews were conducted and all participants had practised ONS (figure 1). A further, willing interviewee unfortunately was not available during the project interval.

Total number of potential participants contacted	13
Number of potential participants contacted due to snowball sampling	4
Number of email replies	11
Number of interviews conducted	5

Figure 1: *Table to show sample statistics and results of snowball sampling.*

2.3 Analysis of interview data

2.3.1 Determination of analysis method

Extensive literature review resulted in thematic analysis (TA) being chosen. TA is a qualitative analytical method that identifies patterns (themes) across the data and summarises findings (figure 2) (figure 3) (Braun and Clarke, 2006; Morse and Field, 1995). See Materials and Methods for further details. TA was chosen as it is flexible and widely used on a diverse range of topics (Fugard and Potts, 2014). TA is independent of theory unlike other major traditions, such as grounded theory which requires detailed technical and theoretical knowledge, so is best for less experienced qualitative researchers (Braun and Clarke, 2006).

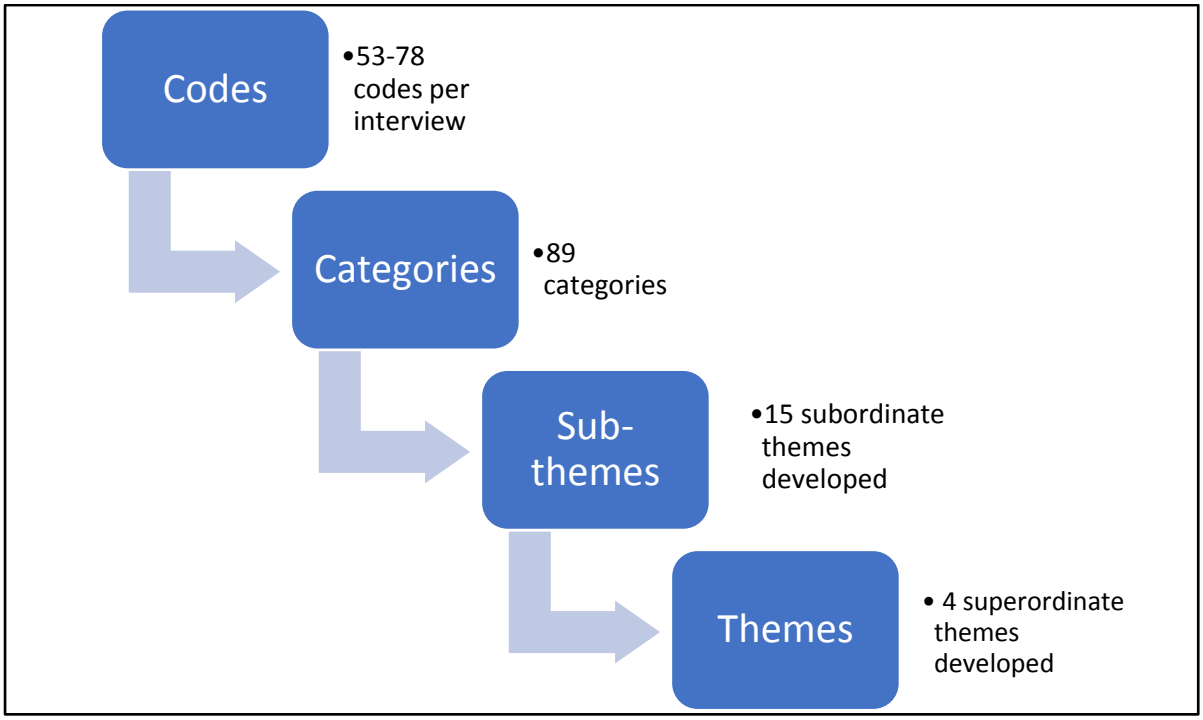

Figure 2: *Flow chart to show the theme development process and results yielded at each part: The interview transcript is coded, identifying any interesting parts of the data. These codes are then grouped into categories by analysing all the codes across the dataset and identifying patterns and relationships between the codes. Categories are grouped into sub-themes based on the significance of the findings and finally sub-themes are refined and form themes individually or by being combined, some sub-themes may be discarded.*

Interview 5	
Data extract	**Coded for**
"I also think if we start using them at an earlier level. So in, you know, undergraduate or even in high schools and train people how to use these notebooks efficiently we could have some really effective teaching and learning in this way"	Start ONS at an early level
	Use ONS for teaching
	Need to train people in ONS

Figure 3*: Table to demonstrate coding of my interview data: This table shows a short extract from interview 5 that has been assigned 3 codes. All the interview transcripts are read and re-read multiple times and initial codes are generated. Codes are raw sections of data that appear interesting to the analyst, based on the aims and objectives of the study, and text extracts may be coded once, many times or not at all (Braun and Clarke, 2006).*

2.3.2 Themes

Four superordinate themes relating to the participants' experiences and attitudes on ONS were yielded. Tables and maps helped the development of a final thematic map, showing superordinate themes and associated subordinate themes (figure 4). Each theme is discussed below with interview quotes to support the findings, see Appendix B for further quotes. Responses are not statistically reported due to the qualitative nature of the study. When text has been removed from the extract for clarity, or to protect identity, it is shown by "(…)".

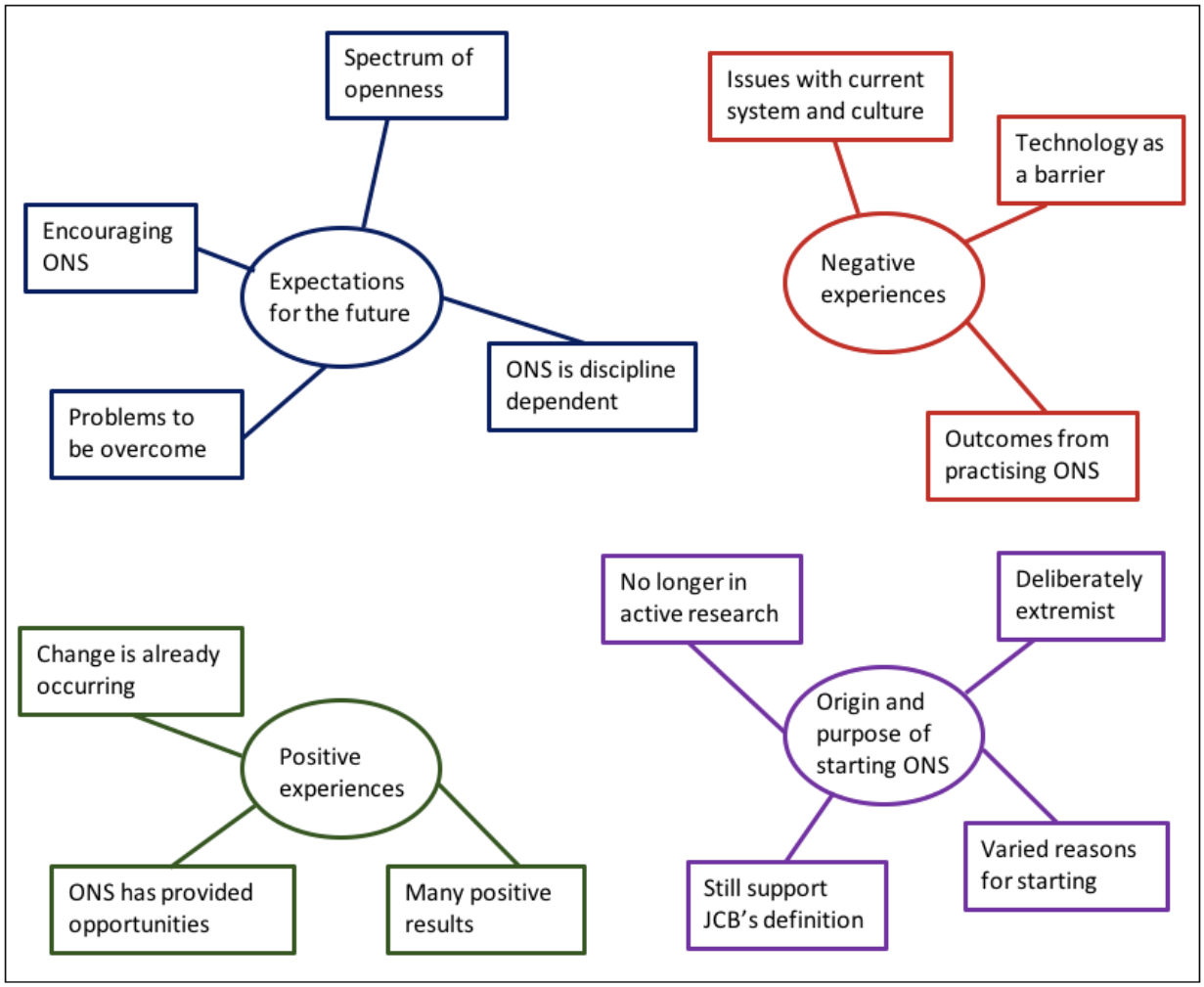

Figure 4: *Final Thematic Map: 4 superordinate themes were developed from the interview data; origin and purpose of starting ONS, positive experiences, negative experiences and expectations for the future. The themes identify patterns within the data and each theme has corresponding subordinate themes within it which are supported by quotes from the interviews.* Key = JCB stands for Jean-Claude Bradley, ellipses represent superordinate themes and rectangles represent subordinate themes.

Origin and purpose of starting ONS

Most participants stated the original definition of ONS. It was mentioned the definition was deliberately strict to avoid confusion with the broader term of Open Science. This translated into participants mentioning that much of the ONS that exists, or has existed, is not true ONS.

> "Other open notebookers don't necessarily release all the data, or all the materials and methods, so it's not actually true open notebooking" [Participant 2]

The majority of participants highlighted that ONS was intentionally extreme and meant to challenge tradition. However, this results in negative opinions of the participants from others. Many participants commented that other people have called them negative names, challenging why they were practising ONS.

> "It was very deliberately an extremist position" [Participant 1]

> "You will get people telling you, you know, that you are, you know, off the wall" [Participant 3]

Participants cited a variety of reasons for starting ONS. Many mentioned they wanted to improve openness, with the opinion that the public deserve access to scientific knowledge, or that it was technically easier to make everything public. Likewise, novelty was revealed as an important aspect of ONS and contributed to starting it.

> "It's better, faster, more inclusive, more innovative" [Participant 3]

> "Research that's funded by taxpayers, by the public, should be visible to the public (…) we owe it to them" [Participant 5]

> "We thought nobody else was really going to be doing it (…) science is all about doing the new, about novelty" [Participant 1]

The majority of participants are no longer in active science research and participants are now involved in areas such as science outreach and education. This is despite all the participants having done undergraduate training or PhD in molecular biology, bio/chemistry or physics.

> "Right now I don't do active research (…) I'm in science education" [Participant 4]

Positive experiences

The majority of participants were keen to point out that practising ONS was positive overall and every participant discussed positive outcomes. Participants highlighted one of the most rewarding results was being able to encourage others. Some participants revealed they had started other projects supporting younger researchers or students in using open notebooks. Many participants stated that ONS leads to positive collaborations.

> "Overall the experience was very positive" [Participant 4]

> "We now have, like a team (…) who are a group of 12 post-doc and PhD students who are all running open notebooks" [Participant 2]

Most participants revealed that ONS has provided them with opportunities. This included increased publicity of their project, the ability to speak at conferences, and acquiring guidance from experts in their field. Equally, many participants mentioned ONS gave them a career boost.

> "I went from being just another not particularly well performing academic in a fairly average UK chemistry department, to having an international profile" [Participant 1]

Finally, participants highlighted the positive change, towards more openness in research, already occurring. Many participants felt there is more sharing today than a few years ago and more people moving away from tradition. Most participants mentioned positive political and policy changes, discussing how encouraging this was. Frequently, participants referred to the ONS technology and said it had improved. Yet, more commonly mentioned was how the technology is still a major barrier today.

> "There's a huge kind of surge in movement of people realising that, you know, closed paywall science is not in the public interest" [Participant 2]

<u>Negative experiences</u>

My findings indicate that technology is a prominent barrier to ONS. Most participants stated it was an issue both when they first started ONS and today. Participants discussed the lack of a universal online platform. One participant stated they changed the style of recording their research so others could reproduce their experiments if they did not have the same experience.

> "The technology is just rubbish" [Participant 1]

> "One of the things that still isn't there yet is the perfect platform" [Participant 5]

Negative outcomes of ONS were discussed by all participants, primarily, how time-consuming ONS is and that it is hard work. The importance of being thorough and uploading information in real-time were reasons cited. It was mentioned the notebook itself requires additional effort, but also by doing ONS you become part of the Open Science community, which further adds to the work. Yet, participants mentioned the extra work pays off as it is easier to write up work, and it is good discipline. When asked whether there was interest from others in their notebook, the majority stated there was not much outside interest. Many discussed negative comments they receive about their notebooks, including others expressing a fear of being scooped and a couple of participants mentioned that they had been scooped. Likewise, participants mentioned their work was not always properly acknowledged or cited. Contrastingly, many participants stated they never had a negative experience with ONS, despite citing at least two of the reasons discussed.

> "It is almost like an extra job" [Participant 2]

> "Did we ever actually have someone randomly come across the lab notebook and offer a comment or solution? No, that never actually happened" [Participant 1]

Every participant negatively mentioned the current culture of how research is shared and how scientists are valued. Participants felt it was vital for their careers to still publish in peer-reviewed journals, yet many expressed frustration with this. Findings indicated a strong opinion, from many participants, about the need for a change in culture.

> "That 'publish or perish' environment is still very prolific" [Participant 2]

> "I think that the culture of about how we share our research and our ideas really has to change" [Participant 5]

<u>Expectations for the future</u>

Regarding the future, all participants want ONS to be encouraged. Yet, many felt it should not be mandated. Most participants do not approve of mandatory practises in general, and would rather that more responsibility be given to researchers. Many stated there should be incentives to encourage ONS, such as prizes. Some participants mentioned it would be beneficial to train and encourage researchers to practise ONS early in their career, or in schools for learning. Many participants felt the need for ONS to be measured to prove its worth and commented on the need for good examples of ONS and organisations which set a precedent.

> "It definitely needs to be encouraged (...) I don't think it should be mandatory" [Participant 5]

> "Until we get to that stage that we have (…) metrics that we put in front of people, it's very difficult to make the case" [Participant 3]

Although participants want ONS to be encouraged, most saw a spectrum of openness in scientific research in the future, as this reflects the different reasons for different levels of sharing. Participants mentioned different levels of openness within open notebooks, together with the broader scale of Open Science. Most participants believed there would be greater sharing of scientific data in general. Some participants wished for everything to be completely open, but were aware this was unlikely.

> "I can almost see having different levels of open notebooks (…) I think the growth might be in terms of a general openness. But also in different levels of openness." [Participant 5]

> "I've always wanted the model to be just completely open, everybody's doing everything completely open. I don't think it'll ever be that way." [Participant 4]

The majority of participants appreciated that different disciplines require different levels of openness. The inability to share sensitive information was cited as the major hindrance, amongst other requirements. However, many participants had the strong opinion that ONS,

or at least Open Access, is vital in certain areas. Specifically, medical research was highlighted and within that, participants stated that rare cancer research would benefit.

> "Different scientists have different requirements" [Participant 5]

> "Why on earth you'd want to keep research to yourself when you are working on some, like, rare form of child brain cancer, is like, frankly beyond me." [Participant 2]

> "Closed access means people die" [Participant 3]

Finally, all participants mentioned issues to be overcome to progress forward in ONS, and openness generally. Most participants expressed worry about no money and resources going towards tools to publish research publically. Many participants mentioned that others fear practising ONS due to scooping, embarrassing themselves online or making a mistake. Finally, participants mentioned that ONS is in fact still quite new and there are few examples for them to follow.

> "You're opening yourself up to criticism" [Participant 5]

> "There's things I have to work out because even people haven't done it before, or (…) the field is different" [Participant 2]

3. Discussion

3.1 Main Findings

Overall, participants felt they had positive experiences of ONS. Interestingly, studies on data sharing have shown similar results. Blumenthal *et al.* (2006) discovered a large proportion of genetic and life scientists only had positive outcomes from sharing data. Also, participants felt ONS was a faster way to conduct science, supporting the claim from existing literature (Gewin, 2013; Stafford, 2010). Likewise, participants stated that they got guidance from experts in their field more readily. Indeed, when an electronic lab notebook was used to investigate the generation of an enantiomer of the drug praziquantel, it was claimed research moved faster due to experts identifying themselves and contributing to the project (Woelfle *et al.*, 2011). Although participants highlighted opportunities resulting from ONS, it is unknown whether this would be experienced by others in the future because findings show novelty was an attractive quality of ONS. Although not well understood, perceived novelty affects the adoption of new technology (Wells *et al.,* 2010). This is associated with high risk, high reward with more emphasis on reward when perceived novelty is high (Agarwal and Prasad, 1998; Wells *et al.,* 2010). Therefore, if ONS' novelty diminishes with time, it could mean less adoption or resulting opportunities.

The intriguing result, where participants insisted they had no negative experiences, could be a social desirability bias, wanting to mask negative characteristics in an attempt to promote ONS (Grimm, 2010). Yet, crucially, my findings reveal that even advocates expressed undesirable consequences from ONS. Participants cited many negative experiences primarily the additional workload, and the overriding need to publish in journals to ensure career advancement. This aligns with a US national survey measuring data withholding in Genetics, where 80% of respondents that deliberately withheld data reasoned it was too much effort, whilst 64% did so to protect publishing prospects (Campbell *et al.,* 2002). My results establish that some of the speculated fears with ONS, such as being scooped, are warranted as some participants admitted to being scooped (Neylon, 2007b; Poynder, 2010; Wald, 2010). Contrasting to my results, publically shared research data has a higher citation rate than unshared data (Piwowar *et al.,* 2007). Therefore, improper acknowledgement reported by participants could be unique to ONS, versus other Open Science practises. This is reinforced because participants pointed to a lack of knowledge on how to cite open notebooks. Interestingly, Nielsen (2009) stated two changes required for extreme openness to work, namely creating excellent online tools and a culture change. Equally, participants believe these changes are fundamental for ONS to work. So, it is perhaps disappointing to hear that issues highlighted in 2009 remain today.

Interestingly, most did not believe ONS should be mandated. Within data reuse and sharing, researchers show concern about mandatory policies because of different challenges faced in different research areas (Van den Eynden *et al.,* 2016). Similarly, even advocates realise ONS may not be appropriate everywhere, and confidential information limits sharing. Existing literature reinforces this, for example, a worldwide survey reported that medical sciences are significantly less willing to share data due to sensitive information (Tenopir *et al.,* 2015). Nonetheless, participants saw a spectrum of openness and strongly felt that ONS is essential in certain areas, exampling rare cancers. Rare cancer research faces difficulties such as lack

of funding, and research not being countries' priority (Miller, 2010; Pillai and Jayasree, 2017). Therefore, ONS could help research publicity and progression.

Unexpectedly, most participants no longer partake in active research and many have projects to encourage others to use ONS, which was highly rewarding. However, it may be a natural progression to work in science outreach. In electronic communities of practise, individuals shared knowledge due to moral obligation, wanting to improve the community and satisfaction of aiding others (Wasko and Faraj, 2000). Finally, my results indirectly show that the prevalence of ONS today is minimal. The extensive literature review, and having to expand my sample from Biological Scientists to other disciplines demonstrates the lack of practitioners. Likewise, snowball sampling results yielded few new contacts despite being beneficial for finding small, rare or interacting populations (Biernacki and Waldorf, 1981; Coleman, 1958; Sadler *et al.,* 2010). The interview data further supports this as participants mentioned they had few examples to follow, most other open notebooks weren't proper ONS, and that ONS needs to be measured to prove its worth to others. However, this deduction has methodological limitations, as with some other areas of my study.

3.2 Limitations and Future Directions

The primary limitation with my study is the small sample size. Five interviews were achieved, whilst literature indicates 6-12 being suitable as TA is about identifying patterns across the data. Likely, individuals did not agree to participate, or reply to emails, as they have no prior relationship with me (Sadler *et al.,* 2010). However, five interviews still yields meaningful results as my data was of high quality, which reduces the required sample size (Guest *et al.,* 2006). Future studies should allow more time, increasing the chance of more individuals agreeing to interview, as a larger sample can discover less common themes (Fugard and Potts, 2015). Regarding methodology, my study is potentially limited by the identification of eligible participants (Biernacki and Waldorf, 1981). Initial participants were identified via literature review, so, sample selection may be biased towards those most written about. However, as ONS is rare and about making everything public, it could be argued that if individuals are hard to find then they are not true advocates. Also, snowball sampling helped counteract this bias. Although prevalence of ONS was inferred, a quantitative measurement would be more conclusive. For example, internet data collection of open notebooks or "altmetrics", which is tracking online events, could be used (Hewson, 2007; Holmberg *et al.,* 2015). A key component to qualitative analysis is identifying factors that may influence results (Meaney *et al.,* 2016). My study did not take into account some defining qualities of each participant, such as career stage. Existing studies have demonstrated that variables such as age and geography can affect attitudes and practices on data sharing (Tenopir *et al.,* 2015; Van den Eynden *et al.,* 2016). Future research should define variables more clearly to identify potential correlations.

For future work in general, it would be beneficial to interview another group. For example, those who have not practised ONS or have but are not advocates, for comparison, as my study cannot determine any significant differences in attitudes between advocates of ONS and others (Padoan *et al.,* 2016). Also, a questionnaire would yield data on the wider opinions surrounding ONS and allow better comparison to surveys on data sharing.

3.3 Implications and Outlook

Despite these limitations, the aims were achieved and my study undoubtedly sheds light on ONS, therefore some conclusions can be drawn. Whilst the future cannot be predicted, my results indicate that ONS will likely remain niche given that those who have practised and support ONS most do not believe it should, or can, be widely adopted. My study showed that practising ONS can be very positive and rewarding, but that does not mean it is feasible for the majority. The low prevalence of ONS, and the negative associations admitted even by its advocates, support this conclusion. Nonetheless, my results indicate a range of openness, including partial-ONS, could be beneficial within scientific research. Partial-ONS involves sharing data after a delay, to subsets of people, or not sharing every experiment (Poynder, 2010). Partial-ONS could prevent some of the negative experiences of true ONS, particularly the time burden and publishing concerns. Therefore, my study reveals some areas to focus on. For example, teaching on how to cite open notebooks correctly, and stricter journal rules, could improve the acknowledgment of open notebooks. Likewise, organisations and institutions should consider encouraging different forms of openness, including ONS, particularly in certain areas. Yet, for informed policies or incentives to be made, further work is warranted as benefits and barriers to sharing data is strongly dependent on research discipline (Fecher *et al.,* 2015; Van den Eynden *et al.,* 2016). Overall, making research data open, particularly in some fields, is still in the early phases. Adoption and development of the practise by the research community is likely to grow (Concordat on Open Research Data, 2016).

3.4 Conclusion

In conclusion, my study addresses a major gap in literature and provides a valuable insight into ONS. Specifically, common experiences and attitudes of advocates of ONS were discovered, providing the first basis of evidence on ONS. Given the results, it is unlikely that true ONS will prevail, but instead will be adapted to suit researchers' needs and partial-ONS may become more widely adopted. With the global movement towards greater transparency in scientific research, it is vital to carefully consider how data is accessed and used. My study helps decipher what sharing infrastructure may be most suitable in scientific research, but, it is challenging to create an efficient and broadly accepted data-sharing framework (Ball *et al.,* 2004). Clearly, there is not one answer for any discipline, organisation or even for an individual researcher. We must achieve a balance and accept that varying levels of secrecy will remain.

4. Materials and Methods

4.1 Experimental protocol

An extensive literature review occurred to determine the best method in line with the aim of this study and to appreciate associated limitations with each potential design. Qualitative semi-structured interviews were chosen, and meetings with an expert in qualitative research confirmed this method was most suitable. Questions are predetermined, but open-ended in nature, and all participants are asked the same questions. To maintain rigour in the interview design process and improve quality, the five-part semi-structured interview development framework by Kallio *et al.* (2016) was followed. The use of preformed questions permits guidance, however questions were deliberately broad, allowing for the participants own response with minimal leading of answers. A list of prompt questions were designed to help the interview flow and encourage more detailed responses if necessary. Participants were also given the chance to add anything at the end of the interview. Two pilot interview tests were conducted to refine the interview quality and ensure relevance of the questions to the study. The first pilot test was by expert assessment from Dr. Niki Vermeulen, who has extensive experience in qualitative interviewing. The second was field testing, which stimulates the real interview, with a University of Edinburgh researcher assuming role of the participant. This allows an estimate of timing, audio and video quality, and practice for the researcher (Kallio *et al.,* 2016). The interview guide *(Appendix A)* was scrutinised after each pilot test. Developing the interview guide rigorously improves the quality and trustworthiness of the method (Kallio *et al.,* 2016).

Skype was used to conduct the interviews, with a quiet, private room to hold the researcher. Use of Skype allowed a broad geographical reach. It was acknowledged that using Skype could impose a bias on the individuals able to participate. However, given the nature of the research question, it was determined that advocates of ONS are likely to be in touch with internet software such as Skype. To reduce this potential bias, a telephone interview was offered as an alternative, although all interviews were conducted via Skype.

4.2 Sampling

The sample was advocates of ONS. The method used was purposeful sampling, specifically mixed-method homogeneity and snowball sampling. Advocates were defined as individuals within the scientific research community who publically supported ONS. Source participants were sought out through the internet via literature review, as by definition advocates make their views public and available. Each participant was contacted via email to request participation in the study. For snowball sampling, source participants were asked via email and verbally to recommend others.

Participant numbers were not limited by theoretical saturation, as often with semi-structured interviews. Theoretical saturation is a concept by Glaser and Strauss (1999) in-line with grounded-theory analysis, which was not used in this study. Due to this, time and resource constraints, and the need to approximate sample size during planning, sample size was determined based on the literature. Due to the size and specificity of my project and

homogeneity of my sample, reviewed literature suggested 6-12 interviews should be sufficient to discover common experiences and perceptions (Fugard and Potts, 2015; Guest *et al.,* 2006; Palinkas *et al.,* 2013). Sample size also depends on the chosen analysis, TA, and Braun and Clarke (2013) suggest 6-10 participants for a small project. Likewise, 'information power' can guide sample size (Malterud *et al.,* 2016). The model is determined by analysing study aim, sample specificity, use of already established theory, dialogue quality and analysis strategy (Malterud *et al.,* 2016). This model was applied and 6-10 interviews, determined during planning, was deemed to have adequate information power.

4.3 Data Collection

Each interview was audio recorded twice, to ensure completeness, and transcribed on the same day. Care was taken to ensure transcripts were complete and generated systematically, with transcripts checked against audiotapes to ensure accuracy. Transcription occurred for all speech between the first question and when the participant had nothing more to add. By transcribing and auditing the audiotapes myself, it allowed data familiarity. Before analysis, each transcript was sent back to the participant, via email, to ensure no misrepresentation. Audiotapes and transcripts were stored securely with identifying details removed.

4.4 Analysis

Analysis of the interview data was carried out manually via TA. TA allows comparison across participants, identifying patterns (themes) across the data and the summarising of findings (Braun and Clarke, 2006; Brooks *et al.,* 2016; Morse and Field, 1995). It involves multiple readings of the transcripts, to ensure data familiarisation, until themes are yielded. First, each transcript is coded, where interesting sections of text are assigned meanings. Then, related codes across transcripts are grouped into categories which then form subordinate themes and superordinate themes. The analysis was conducted through a realist method which reports motivations and experience of participants and is most common (Braun and Clarke, 2006). Themes were identified through a theoretical approach as this involves engagement with the literature before analysis (Braun and Clarke, 2006). Some argue that early literature review can narrow findings, leading to a too focused, or influenced view that may neglect some vital aspects in the data, however others state that it can improve analysis as more sensitive features can be identified (Braun and Clarke, 2006; Tuckett, 2005). Themes were identified on a semantic level, as a TA focusing on semantic themes is often the realist method (Braun and Clarke, 2006). The 15 point checklist for the TA process, by Braun and Clarke (2006), was followed to ensure a thorough analysis.

4.5 Ethics

All participants were contacted via email and sent an interview information sheet *(Appendix C)* and consent form *(Appendix D)*. Written and verbal consent was obtained from all participants with their anonymity guaranteed, unless otherwise requested. Ethical approval was obtained from the University of Edinburgh Biological Sciences department.

5. References

Agarwal R, & Prasad J (1998) The antecedents and consequents of user perceptions in information technology adoption. *Decision Support Systems*, 22: 15–29

Ball C, Sherlock G & Brazma A (2004) Funding high-throughput data sharing. *Nature Biotechnology* 22: 1179-1183

Bergman O (2014) Open Notebook Science – the Carl Boettiger Interview. *Crastina.se* Available at: http://crastina.se/open-notebook-science-the-carl-boettiger-interview-2/ [Accessed May 11, 2018]

Biernacki P & Waldorf D (1981) Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research* 10: 141-163

Blumenthal D, Campbell E, Gokhale M, Yucel R, Clarridge B, Hilgartner S & Holtzman N (2006) Data Withholding in Genetics and the Other Life Sciences: Prevalences and Predictors. *Academic Medicine* 81: 137-145

Bradley J-C, Neylon C, Guha R, Williams A, Hooker B, Lang A, Friesen B, Bohinski T, Bulger D, Federici M, Hale J, Mancinelli J, Mirza K, Moritz M, Rein D, Tchakounte C & Truong H (2010) Open Notebook Science Challenge: Solubilities of Organic Compounds in Organic Solvents. Available from Nature Precedings <http://dx.doi.org/10.1038/npre.2010.4243.3> [Accessed April 30, 2018]

Braun V & Clarke V (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3: 77-101

Braun V & Clarke V(2013). Successful qualitative research: A practical guide for beginners. London: Sage.

Brooks S, Gerada C & Chalder T (2016) The specific needs of doctors with mental health problems: qualitative analysis of doctor-patients' experiences with the Practitioner Health Programme. *Journal of Mental Health* 26: 161-166

Campbell E, Clarridge B, Gokhale M, Birenbaum L, Hilgartner S, Holtzman N & Blumenthal D (2002) Data Withholding in Academic Genetics. *JAMA* 287: 473

Clinio A & Albagli S (2017) Open notebook science as an emerging epistemic culture within the Open Science movement. *Revue française des sciences de l'information et de la communication*

Coleman J (1958) Relational Analysis: The Study of Social Organizations with Survey Methods. *Human Organization* 17: 28-36

Concordat on Open Research Data (2016) Available at: http://www.hefce.ac.uk/news/newsarchive/2016/Name,109264,en.html [Accessed April 30, 2018]

De Roure D, Goble C, Aleksejevs S, Bechhofer S, Bhagat J, Cruickshank D, Fisher P, Hull D, Michaelides D, Newman D, Procter R, Lin Y & Poschen M (2010) Towards open science: the myExperiment approach. *Concurrency and Computation: Practice and Experience* 22: 2335-2353

DiCicco-Bloom B & Crabtree B (2006) The qualitative research interview. *Medical Education* 40: 314-321

Etikan I (2016) Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics* 5: 1

Fecher B, Friesike S & Hebing M (2015) What Drives Academic Data Sharing?. *PLOS ONE* 10: e0118053

Franco A, Malhotra N & Simonovits G (2014) Publication bias in the social sciences: Unlocking the file drawer. *Science* 345: 1502-1505

Friesike S & Schildhauer T (2014) Open Science: Many Good Resolutions, Very Few Incentives, Yet. *Incentives and Performance*: 277-289

Fugard A & Potts H (2015) Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology* 18: 669-684

Gewin V (2013) Turning point: Carl Boettiger. *Nature* 493: 711-711 Available at: https://www.nature.com/naturejobs/science/articles/10.1038/nj7434-711a [Accessed May 11, 2018]

Glaser B & Strauss A (1999). Discovery of Grounded Theory. New York: Routledge.

Grand A, Wilkinson C, Bultitude K & Winfield A (2016) Mapping the hinterland: Data issues in open science. *Public Understanding of Science* 25: 88-103

Grimm P (2010) Social Desirability Bias. *Wiley International Encyclopedia of Marketing*

Guest G, Bunce A & Johnson L (2006) How Many Interviews Are Enough?. *Field Methods* 18: 59-82

Hewson C (2007) Gathering data on the Internet. *The Oxford handbook of Internet psychology*: 406-428.

Holmberg K, Didegah F, Bowman T & Kortelainen T (2015). Measuring the societal impact of open science–Presentation of a research project. *Informaatiotutkimus 34*.

Insel T (2013) Post by Former NIMH Director Thomas Insel: Open Data. *National Institute of Mental Health* Available at: https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/open-data.shtml [Accessed May 11, 2018]

Kallio H, Pietilä A, Johnson M & Kangasniemi M (2016) Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing* 72: 2954-2965

Malterud K, Siersma V & Guassora A (2016) Sample Size in Qualitative Interview Studies. *Qualitative Health Research* 26: 1753-1760

McNutt M (2014) Reproducibility. *Science* 343: 229-229

Meaney S, Everard C, Gallagher S & O'Donoghue K (2016) Parents' concerns about future pregnancy after stillbirth: a qualitative study. *Health Expectations* 20: 555-562

Miller R (2010) Problems in rare tumor study: a call for papers. *Rare Tumors* 2: 46-47

Morse J & Field P (1995). Qualitative research methods for health professionals. Thousand Oaks, CA: Sage.

Murray-Rust P (2014) Jean-Claude Bradley: Hero of Open Notebook Science; it must become the central way of doing science. *Petermr's Blog* Available at: https://blogs.ch.cam.ac.uk/pmr/2014/05/19/jean-claude-bradley-hero-of-open-notebook-science-it-must-become-the-central-way-of-doing-science/ [Accessed May 11, 2018]

Nielsen M (2009) Doing science in the open. *Physics World* 22: 30-35 Available at: http://iopscience.iop.org/article/10.1088/2058-7058/22/05/38/meta [Accessed May 11, 2018]

Neylon C (2007a) How best to do the open notebook thing…a nice specific example. *Science in the Open* Available at: http://cameronneylon.net/blog/how-best-to-do-the-open-notebook-thinga-nice-specific-example/ [Accessed May 11, 2018]

Neylon C (2007b) Getting Scooped... *Science in the Open* Available at: http://cameronneylon.net/blog/increasing-the-persistence-of-online-open-notebooks/ [Accessed May 11, 2018]

Nosek B, Alter G, Banks G, Borsboom D, Bowman S, Breckler S, Buck S, Chambers C, Chin G, Christensen G, Contestabile M, Dafoe A, Eich E, Freese J, Glennerster R, Goroff D, Green D, Hesse B, Humphreys M & Ishiyama J et al (2015) Promoting an open research culture. *Science* 348: 1422-1425

Padoan C, Garcia L, Rodrigues A, Patusco L, Atz M, Kapczinski F, Goldim J & Magalhães P (2016) "Why throw away something useful?": Attitudes and opinions of people treated for bipolar disorder and their relatives on organ and tissue donation. *Cell and Tissue Banking* 18: 105-117

Palinkas L, Horwitz S, Green C, Wisdom J, Duan N & Hoagwood K (2013) Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research* 42: 533-544

Pillai R & Jayasree K (2017) Rare cancers: Challenges & issues. *Indian Journal of Medical Research* 145: 17

Piwowar H & Chapman W (2008) A review of journal policies for sharing research data. *ELPUB* 1-14.

Piwowar H, Day R & Fridsma D (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2: e308

Poynder R (2010) Interview With Jean-Claude Bradley The Impact of Open Notebook Science. *Information Today* Available at: http://www.infotoday.com/IT/sep10/Poynder.shtml [Accessed March 9, 2018]

Randles B, Pasquetto I, Golshan M & Borgman C (2017) Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*

Sadler G, Lee H, Lim R & Fullerton J (2010) Research Article: Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing & Health Sciences* 12: 369-374

Stafford N (2010) Science in the digital age. *Nature* 467: S19-S21

Tenopir C, Dalton E, Allard S, Frame M, Pjesivac I, Birch B, Pollock D & Dorsett K (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* 10: e0134826

The Royal Society (2012) Science as an open enterprise Available at: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/ [Accessed March 7, 2018]

Tuckett A (2005) Applying thematic analysis theory to practice: A researcher's experience. *Contemporary Nurse* 19: 75-87

Van den Eynden V, Knight G, Vlad A, Radler B, Tenopir C, Leon D, Manista F, Whitworth J & Corti L (2016) Towards Open Research: Practices, experiences, barriers and opportunities. *Wellcome Trust.*

Wald C (2010) Scientists Embrace Openness. *Science*

Wasko M & Faraj S (2000) “It is what one does”: why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems* 9: 155-173

Wells J, Campbell D, Valacich J & Featherman M (2010) The Effect of Perceived Novelty on the Adoption of Information Technology Innovations: A Risk/Reward Perspective. *Decision Sciences* 41: 813-843

Whiting L (2008) Semi-structured interviews: guidance for novice researchers. *Nursing Standard* 22: 35-40

Woelfle M, Olliaro P & Todd M (2011) Open science is a research accelerator. *Nature Chemistry* 3: 745-748

6. Acknowledgements

Primarily, an enormous thank you to Professor Andrew Millar for introducing me to the fascinating topic of ONS and for his continual support and enthusiasm. I am also extremely grateful to all those who participated in my study, for making the interviews enjoyable and interesting. I thank Dr.Niki Vermeulen and Dr.Ros Attenborough for their advice and guidance, and Dr. Clare Blackburn for her pilot test and feedback. I also thank Argyris, Eilidh and Tomasz for their company, encouragement and suggestions throughout my project.

7. Appendices

7.1 Appendix A: Interview Guide

Background questions

1. What field of scientific research would you define yourself to be a part of?
2. How did you become a part of that field?
3. How would you define Open Notebook Science?
4. Do you currently have an active open notebook?
5. If yes – Is it available and free to the public? If not, who is it open to?
   If no – When did you last have an open notebook? Why do you not currently have one?

Experience practicing ONS

1. When did you first hear about ONS?
2. Why did you start an open notebook?
3. How would you describe your experience in practicing ONS? (prompt e.g. positive, negative, any examples)
4. Do you think your experiences have changed since you started? If so, how?
5. Did you face any barriers when you first started ONS? If so, what?
6. Are the motivations and barriers you face today, different from when you first started? If so, how?
7. Do others express interest in your work? (prompt e.g. positive or negative, describe what type of interest)
8. Has this changed over your time practicing ONS?

Opinions on the future

1. Do you think ONS needs to be encouraged further? If so, what incentives do you think would encourage it?
2. What do you think the future holds for ONS?
3. What do you think is the most realistic model for openness within scientific research? (prompt e.g. open data, open access, different forms of ONS)

Snowball sampling

1. Do you know anybody who uses ONS today?
2. Do you know anybody who used to use ONS, but has since stopped?
3. If so, would it be possible for you to put me in touch with them?

Ask if the participant would like to talk about anything else they may feel is important which has not been covered by these questions and thank them for their time.

Prompt questions

- Could you tell me more about this?
- What exactly do you mean by xxx?
- Could you give me an example?

7.2 Appendix B: Interview Extracts

Superordinate Themes	Subordinate Themes	Supporting Interview Extracts
Origin and purpose of starting ONS	Still support JCB's definition	"I'm very glad that he did this because the term is very clear, nobody confuses it. Whereas open science can mean whatever anybody wants it to mean" [participant 3]
		"I would probably still go back to the definitions that Jean-Claude Bradley" [participant 1]
	Deliberately extremist	"So open notebook science is a public statement" [participant 3]
		"We were interested in kind of challenging those norms" [participant 2]
	Varied reasons for starting	"Just making it all available was driven, at least in significant part, by security problems and issues with data management" [participant 1]
		"I'd grown, was growing increasingly frustrated with the lack of openness" [participant 5]
	No longer in active research	"So I am doing less work in the lab than ever (…) So, well my current title is that I'm a lecturer in chemistry education and outreach" [participant 5]
Positive experiences	Many positive outcomes	"What we wanted to do was document everything that we did so that even if we weren't successful, like, other people would be able to pick up where we left off" [participant 2]
		"So I have an undergraduate class who do real research in their first year and they use open notebooks" [participant 5]
	ONS has provided opportunities	"It's afforded me a large number of opportunities doing things such as, well I've been able to, like, go, been invited to give loads of talks in loads of different places" [participant 2]
	Change is already occurring	"I think something that's shifted, again 10 years ago, very few people were sharing anything, beyond very specific data" [participant 1]
		"I think probably an increased interest and an increased acceptance that it's a genuine method" [participant 5]
Negative experiences	Technology as a barrier	"The biggest objective barrier, without question, was the lack of any generic technology" [participant 3]
		"I had to start adapting the way I write for, for other audiences" [participant 4]
	Outcomes from practising ONS	"Some people think it's a bit of a fad" [participant 5]
		"But we got scooped on a couple of things" [participant 1]
		"I've always received like a, kind of a push-back" [participant 4]
	Issues with current system and culture	"There's this notion of what counts, still remains a really big barrier. You have to be in the first place that counts" [participant 1]
		"the problem is that we still are focused on an impact driven publication system in order to rank who we decide are the best scientists and that is incredibly frustrating" [participant 2]
Expectations for the future	Encouraging ONS	"I also think if we start using them at an earlier level. So in, you know, undergraduate or even in high schools and train people how to use these notebooks efficiently we could have some really effective teaching and learning in this way" [participant 5]
	Spectrum of Openness	"I guess that I always try to kind of present it as this spectrum of you know, how comfortable are you presenting stuff online.' [participant 4]
	ONS is discipline dependent	"If you have things like sensitive data, human data, indigenous rights, rare species, etcetera, and so on. Then there's a higher authority which says you can't publish this immediately because there are serious downstream problems" [participant 3]
		"There are very valid reasons why some open practices just aren't applicable in different areas of science research" [Participant 2]
	Problems to be overcome	"My main problem in science is actually getting anybody to listen to what I'm doing" [participant 3]
		"We need to be more efficient with the budgets that we have for research because it's so competitive" [participant 5]

Table 1: Interview Extracts to support the Results: This table accompanies the Results section to give extra support to the themes yielded from the data. Please note this is not the raw data or all of the interview extracts that support each theme and subordinate theme.

7.3 Appendix C: Interview Information Sheet

Interview information: Qualitative interview study on Open Notebook Science

What will the interview involve?
The interview will involve questions regarding your experiences with Open Notebook Science, how you became involved, and your opinions surrounding it. The interview will last approximately 45 minutes and will be, with your permission, audio recorded. I am interested in your side and, therefore, there are no correct or incorrect answers.

What are the benefits?
The data from your interview will be the centre part of my Honours project on Open Notebook Science which will not only help to raise awareness on Open Notebook Science, but also help form models for openness in science and encourage further research into Open Notebook Science.

Who is doing the interviews?
The researcher of the entire study is myself, Amelia Hunt, and I will also be conducting the interviews. I am an Honours student in Biological Sciences (Genetics) at The University of Edinburgh and my supervisor is Professor Andrew Millar.

Will I be anonymous?
By default you will remain anonymous, unless you specifically request otherwise. In any records or references which refer to your interview, a pseudonym or number will be used instead of your real name. Any other details (e.g. email or job title) that may identify you will also not be included. Your consent will be required (written, via the consent form, and verbal, at the start of the interview) before any data is recorded or archived. You may request to waive your anonymity if you wish.

What will happen to the interview data (audio recordings and transcripts)?
The audio recordings and transcript will be kept securely by myself for analysis, following your consent. The audio recording and transcript will only be accessible to myself and my project supervisor, Professor Andrew Millar. You will have the opportunity to view the transcript of the interview and give any additional feedback or comments. With your permission, I will keep your anonymised interview transcript for the duration of my research project and it will be deleted when it is not needed any longer.

Are there any risks?
This research is not expected to cover any sensitive topics and will occur in a safe environment of your choice, as the interviews will be via Skype or telephone. Therefore, it is considered very low risk. In order to make sure you are not misrepresented, you will be provided with a copy of the interview transcript to check over.

Can I withdraw from the interview?
You have the right to withdraw from the interview process at any time. You also have the right to retract any comments or data after the interview as well.

If you have any questions please contact me (Amelia Hunt) at [email address]

7.4 Appendix D: Consent Form

Consent form: Qualitative Interview study on Open Notebook Science

Please indicate **Yes** or **No** in the boxes below to indicate whether you agree and have understood each of the statements.

I have read and understood the Interview information sheet, and I understand my role as a participant	
I understand that my participation is voluntary and that I may withdraw from the interview at any time	
I understand that I will be anonymous in any information reported from the interview, unless I request otherwise	
I agree to audio recording of the interview for the purpose of transcription. The recording will only be accessible to the researcher (Amelia Hunt) and her project supervisor (Prof. Andrew Millar)	
I understand that after I have been sent the interview transcript I will be offered a choice about the archiving of data from my interview. No data will be archived without my consent	
I give permission for the researcher to securely retain an anonymized interview transcript for the purpose of the research project. It will be deleted when it is no longer needed for this purpose. The transcript will only be accessible to the researcher (Amelia Hunt) and her supervisor (Prof. Andrew Millar)	
I agree to participate in this study	

Name of participant:

Signature of participant:

Date: